PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/47, A61K 38/17, C12N 5/10, C12Q 1/68

A2

(11) International Publication Number: WO 98/45437

(43) International Publication Date: 15 October 1998 (15.10.98)

(21) International Application Number: PCT/US98/06956

(22) International Filing Date: 10 April 1998 (10.04.98)

(30) Priority Data:
08/837,312 10 April 1997 (10.04.97) US

(71) Applicant: GENETICS INSTITUTE, INC. [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(72) Inventors: JACOBS, Kenneth; 151 Beaumont Avenue, Newton, MA 02160 (US). MCCOY, John, M.; 56 Howard Street, Reading, MA 01867 (US). LAVALLIE, Edward, R.; 113 Ann Lee Road, Harvard, MA 01451 (US). RACIE, Lisa, A.; 124 School Street, Acton, MA 01720 (US). MERBERG, David; 2 Orchard Drive, Acton, MA 01720 (US). TREACY, Maurice; 93 Walcott Road, Chestnut Hill, MA 02167 (US). SPAULDING, Vikki; 11 Meadowbank Road, Billerica, MA 01821 (US). AGOSTINO, Michael, J.; 26 Wolcott Avenue, Andover, MA 01810 (US).

(74) Agent: SPRUNGER, Suzanne, A.; Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SECRETED EXPRESSED SEQUENCE TAGS (sESTs)

(57) Abstract

Secreted expressed sequence tags (sESTs) isolated from a variety of human tissue sources are provided.

‡: The hybrid length is that anticipated for the hybridized region(s) of the hybridizing polynucleotides. When hybridizing a polynucleotide to a target polynucleotide of unknown sequence, the hybrid length is assumed to be that of the hybridizing polynucleotide. When polynucleotides of known sequence are hybridized, the hybrid length can be determined by aligning the sequences of the polynucleotides and identifying the region or regions of optimal sequence complementarity.

†: SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete.

*T_B - T_R: The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T_m(°C) = 81.5 + 16.6($\log_{10}$ [Na^+]) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and [Na^+] is the concentration of sodium ions in the hybridization buffer ([Na^+] for 1xSSC = 0.165 M).

Additional examples of stringency conditions for polynucleotide hybridization are provided in Sambrook, J., E.F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, chapters 9 and 11, and *Current Protocols in Molecular Biology*, 1995, F.M. Ausubel et al., eds., John Wiley & Sons, Inc., sections 2.10 and 6.3-6.4, incorporated herein by reference.

Preferably, such hybridizing polynucleotides have at least 70% sequence identity (more preferably, at least 80% identity; most preferably at least 90% or 95% identity) with the polynucleotide of the present invention to which they hybridize, where sequence identity is determined by comparing the sequences of the hybridizing polynucleotides when aligned so as to maximize overlap and identity while minimizing sequence gaps. The isolated polynucleotide encoding the protein of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman *et al.*, Nucleic Acids Res. 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, Methods in Enzymology 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control

sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

A number of types of cells may act as suitable host cells for expression of the protein. Mammalian host cells include, for example, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from in vitro culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Kluyveromyces* strains, *Candida,* or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium,* or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, California, U.S.A. (the MaxBac® kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (i.e., from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl® or Cibacrom blue 3GA

USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified below. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or by administration or use of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA).

Research Uses and Utilities

The polynucleotides provided by the present invention can be used by the research community for various purposes. The primary use of polynucleotides of the invention which are sESTs is as porbes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond (i.e., is a longer polynucleotide sequence of which substantially the entire sEST is a fragment in the case of a full-length cDNA, or which encodes the sEST in the case of a genomic DNA molecule) to such sESTs. Techniques for use of such sequences as probes for larger cDNAs or genomic molecules are well known in the art.

The polynucleotides can also be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify

```
GGCAGTCTAA AACCCACATC TACCATTTCC ACAAGCCCTC CCTTGATCCA TAGCTTTGTT     420
TCTAAAGTGC CTTGGAATGC ACCTATAGCA GATGAAGATC TTTTGCCCAT CTCAGCACAT     480
CCCAATGCTA CACCTGCTCT GTCTTCAGAA AACTTCACTT GGTCTTTGGT CAATGACACC     540
GTGAAAACTC GTGATAACAG TTCCATTACA GTTAGCATCC TCTCTTCAGA ACCAACTTCT     600
CCATCTGTGA CCCCCTTGAT AGTGGAACCA AGTGGATGGC TTACCACAAA CAGTGATAGC     660
TTCACTGGGT TTACCCCTTA TCAAGAAAAA ACAACTCTAC CTACC                    705
```

(2) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 729 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

```
GAATTCGGCC AAAGAGGCCT ACCTTACTTG AGTCCACAGG CAAGGCCCAA TAATGCATAT      60
ACTGCCATGT CAGATTCCTA CTTACCCAGT TACTACAGTC CCTCCATTGG CTTCTCCTAT     120
TCTTTGGGTG AAGCTGCTTG GTCTACGGGG GGTGACACAG CCATGCCCTA CTTAACTTCT     180
TATGGACAGC TGAGCAACGG AGAGCCCCAC TTCCTACCAG ATGCAATGTT TGGGCAACCA     240
GGAGCCCTAG GTAGCACTCC ATTTCTTGGT CAGCATGGTT TTAATTTCTT TCCCAGTGGG     300
ATTGACTTCT CAGCATGGGG AAATAACAGT TCTCAGGGAC AGTCTACTCA GAGCTCTGGA     360
TATAGTAGCA ATTATGCTTA TGCACCTAGC TCCTTAGGTG GAGCCATGAT TGATGGACAG     420
TCAGCTTTTG CCAATGAGAC CCTCAATAAG GCTCCTGGCA TGAATACTAT AGACCAAGGG     480
ATGGCAGCAC TGAAGTTGGG TAGCACAGAA GTTGCAAGCA ATGTTCCAAA AGTTGTAGGT     540
TCTGCTGTTG GTAGCGGGTC CATTACTAGT AACATCGTGG CTTCCAATAG TTTGCCTCCA     600
GCCACCATTG CTCCTCCAAA ACCAGCATCT TGGGCTGATA TTGCTAGCAA GCCTGCAAAA     660
CAGCAACCTA AACTGAAGAC CAAGAATGGC ATTGCAGGGT CAAGTCTTCC GCCACCCCCA     720
ACACTCGAG                                                            729
```

(2) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 686 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

```
GAATTCGGCC AAAGAGGCCT ACTACCATGT CCTCTTGGAG CAGACAGCGA CCAAAAAGCC      60
CAGGGGGCAT TCAACCCCAT GTTTCTAGAA CTCTGTTCCT GCTGCTGCTG TTGGCAGCCT     120
CAGCCTGGGG GGTCACCCTG AGCCCCAAAG ACTGCCAGGT GTTCCGCTCA GACCATGGCA     180
GCTCCATCTC CTGTCAACCA CCTGCCGAAA TCCCCGGCTA CCTGCCAGCC GACACCGTGC     240
ACCTGGCCGT GGAATTCTTC AACCTGACCC ACCTGCCAGC CAACCTCCTC CAGGGCGCCT     300
CTAAGCTCCA AGAATTGCAC CTCTCCAGCA ATGGGCTGGA AAGCCTCTCG CCCGAATTCC     360
TGCGGCCAGT GCCGCAGCTG AGGGTGCTGG ATCTAACCCG AAACGCCCTG ACCGGGCTGC     420
CCTCGGGCCT CTTCCAGGCC TCAGCCACCC TGGACACCCT GGTATTGAAA GAAAACCAGC     480
TGGAGGTCCT GGAGGTCTCG TGGCTACACG GCCTGAAAGC TCTGGGGCAT CTGGACCTGT     540
CTGGGAACCG CCTCCGGAAA CTGCCCCCCG GGCTGCTGGC CAACTTCACC CTCCTGCGCA     600
CCCTTGACCT TGGGGAGAAC CAGTTGGAGA CCTTGCCACC TGACCTCCTG AGGGGTCCGC     660
TGCAATTAGA ACGGCACATT CTCGAG                                         686
```

(2) INFORMATION FOR SEQ ID NO:263: